

Д. И. Барсуков, В. В. Горбань, А. Н. Савиных
(Москва, ФУМО ИБ). Выявление микромимических реакций: опыт обучения нейросетевых алгоритмов на малых выборках.

УДК 004.85

DOI https://doi.org/10.52513/08698325_2024_31_1_1

Резюме: В работе представлен опыт обучения нейросетевых алгоритмов для решения задачи классификации видеоданных на основе выявления микромимических реакций.

Ключевые слова: компьютерное зрение, машинное обучение, искусственные нейронные сети.

Количественная калибровка определения степени выраженности микромимических реакций и выявления лиц с похожими проявлениями реакций в видеоряде, на котором демонстрируется нейтральный и стимульные материалы, основано на детектировании на кадрах глаз человека и анализа среднего соотношения сторон размеров обоих глаз (далее обозначается символом D), которое вычисляется по следующей формуле:

$$D = \frac{\|P_2^l - P_6^l\| + \|P_3^l - P_5^l\|}{4 \cdot \|P_1^l - P_4^l\|} + \frac{\|P_2^r - P_6^r\| + \|P_3^r - P_5^r\|}{4 \cdot \|P_1^r - P_4^r\|},$$

где P_i^l — точки контура левого глаза, P_i^r — точки контура правого глаза, $\|\cdot\|$ — L_2 -норма.

Детектирование контура является классической задачей компьютерного зрения и решается с помощью таких моделей как *SPIGA*, *FAN3D*, *STAR*, моделями *OpenCV*.

После кадрирования видео с постоянной частотой 30 кадров в секунду и покадрового расчета в соответствие видео ставится пара последовательностей $(\tilde{D}_1, \tilde{D}_2)$, где $\tilde{D}_1$ — последовательность D кадров первой части видео с нейтральным материалом), $\tilde{D}_2$ — последовательность D кадров второй части видео со стимульным материалом.

При исследовании пар последовательностей для обеих групп респондентов в ходе обучения модели, установлено, что при $D < 0,22$ на видео происходят специфические движения глаз, что интерпретируется как характерные микромимические реакции на предлагаемый респонденту видеоряд. При этом средняя длительность реакции составляет около 0,2 секунд. Таким образом, для пары последовательностей $(\tilde{D}_1, \tilde{D}_2)$ осуществляется извлечение подпоследовательностей длительностью не менее 0,2 секунд.

Для решения задач оценки выраженности реакций и поиска похожих респондентов был проведен следующий анализ. Для значений обеих последовательностей респондентов первой и второй группы осуществлялось построение эмпирической функции распределения (ЭФР). На Рис. 1 желтым и коричневым цветом изображены ЭФР первой и второй последовательности первой группы соответственно, зеленым и красным цветом — ЭФР первой и второй последовательности второй группы соответственно.

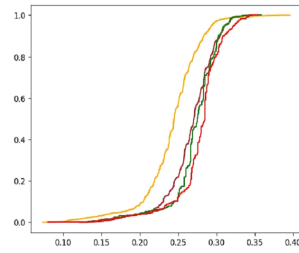


Рис. 1. ЭФР метрики для первой и второй групп

Из рисунка видно, что в первой группе присутствует смещение ЭФР, что говорит о более широком раскрытии глаз на 2 части видео, что может отражать различия в реакциях респондентов на оказываемое влияние второй части видео. Во второй группе раскрытие глаз на протяжении всего видеоряда изменяется слабо, что говорит об отсутствии различных реакций. Таким образом, в качестве признака для оценки выраженности реакций используются доли кадров, на которых происходит реакция, относительно всех кадров каждой части видео.

Для определения микромимических реакций как специфичного изменения геометрии глаз во времени последовательности могут рассматриваться как временные ряды, для которых может быть вычислен разность первого порядка. На Рис. 2 изображены гистограммы разностей для первой части видео респондентов первой группы (зеленым цветом) и второй группы (красным цветом).

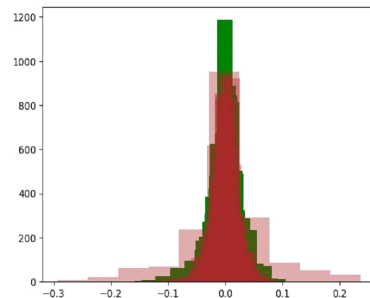


Рис. 2. Гистограммы разностей для первой части

На Рис. 3 изображены гистограммы разностей для второй части видео респондентов первой группы (зеленым цветом) и второй группы (красным цветом).

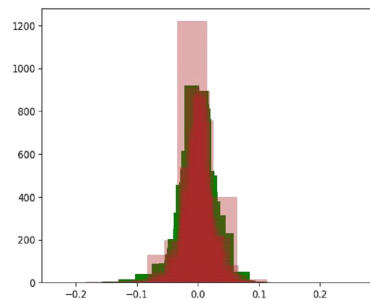


Рис. 3. Гистограммы разностей для второй части

Из рисунков видно, что в основном изменения в видео во второй группе обладают меньшей дисперсией, относительно изменений в видео первой группы. Это свидетель-

ствует о том, что во второй группе реакции имеют более резкий характер (что интерпретируется как микромимическая реакция), в сравнении с респондентами первой группы.

По результатам проведенного исследования, в качестве признакового пространства для решения задач количественной калибровки степени выраженности реакций и поиска респондентов с похожими реакциями предлагается использовать доли кадров, на которых происходят реакции, относительно всех кадров каждой из частей видео, а также 11 квантилей (выбранных равномерно) выборочного распределения первых разностей последовательностей первой и второй частей видео.

Решение первой задачи (степень выраженности) достигается путем обучения модели, имеющей следующую архитектуру (Рис. 4, слева). Модель предсказывает вероятность того, что входной вектор соответствует респондентам второй группы.

<code>nn.Linear(2 + 11 * 2, 512)</code>	<code>nn.Linear(2 * (2 + 11 * 2), 128)</code>
<code>nn.ReLU()</code>	<code>nn.ReLU()</code>
<code>nn.Linear(512, 512)</code>	<code>nn.Linear(128, 512)</code>
<code>nn.ReLU()</code>	<code>nn.ReLU()</code>
<code>nn.Dropout(0.1)</code>	<code>nn.Linear(512, 512)</code>
<code>nn.Linear(512, 512)</code>	<code>nn.Dropout(0.1)</code>
<code>nn.ReLU()</code>	<code>nn.ReLU()</code>
<code>nn.Dropout(0.1)</code>	<code>nn.Linear(512, 512)</code>
<code>nn.Linear(512, 512)</code>	<code>nn.ReLU()</code>
<code>nn.Dropout(0.1)</code>	<code>nn.Dropout(0.1)</code>
<code>nn.ReLU()</code>	<code>nn.Linear(512, 2)</code>
<code>nn.Linear(512, 2)</code>	<code>nn.Linear(512, 2)</code>
<code>nn.Softmax()</code>	

Рис. 4. Архитектуры моделей

Решение второй задачи выполнялось путем обучения модели, имеющей архитектуру, представленную на Рис. 4 справа.

Путем сравнения ответа модели с нулем осуществляется предсказание того, что два поданных на вход модели вектора, которые соответствуют двум разным видео, относятся к одной и той же группе респондентов.

Качество моделей оценивалось с помощью *accuracy*. Для первой модели *accuracy* = 0,8142, для второй модели *accuracy* = 0,9514.

Поступила в редакцию
22.XII.2024

UDC 004.85

DOI https://doi.org/10.52513/08698325_2024_31_1_1

D. I. Barsukov, V. V. Gorban, A. N. Savinykh (Moscow, FEMA IS). Detection of micromimical reactions: experience of training neural network algorithms on small samples

Abstract: The paper presents the experience of learning neural network algorithms to solve the problem of classifying video data based on the detection of micromimical reactions.

Keywords: computer vision, machine learning, artificial neural networks.